Effect of Vitamin C on Tooth Growth in Guinea Pigs

Liubov Gryaznova

November 22, 2015

Overview

In this project we analyze the ToothGrowth data in the R datasets package. The data set provides the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (coded as VC).

Basic Exploratory Data Analyses

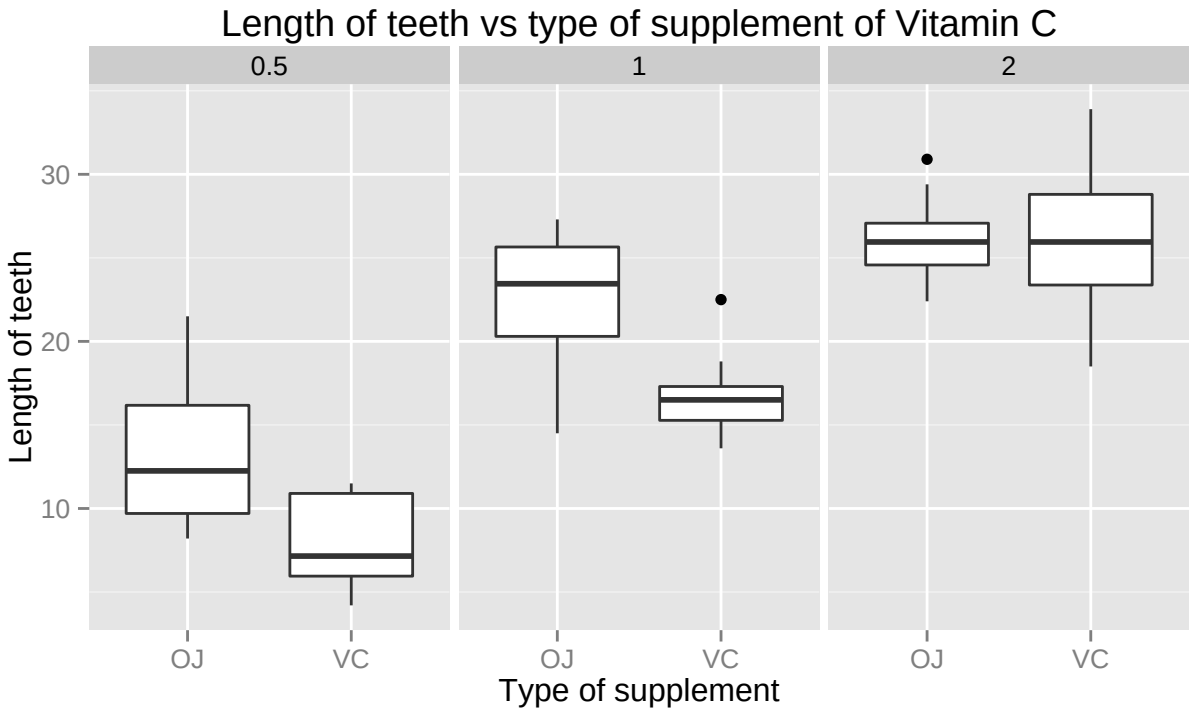
We load the data set and look at its structure to work out further strategy.

```
data(ToothGrowth); str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

We have only two factors which may affect tooth growth in this experiment, namely delivery method and dose of Vitamin C. So we make panel plots to see if there may be correlation. We use **ggplot2** package for our convenience.

```
library(ggplot2)
ggplot(ToothGrowth, aes(x = supp, y = len)) + geom_boxplot() + facet_grid(. ~ dose) +
    ggtitle("Length of teeth vs type of supplement of Vitamin C") +
    ylab("Length of teeth") + xlab("Type of supplement")
```



The plot demonstrates the length of teeth versus the type of supplement of Vitamin C ('OJ' stands for 'orange juice' and 'VC' stands for 'ascorbic acid') given three dose levels. We see that higher doses correlate with greater length. As concerns the type of supplement, orange juice correlates with greater length given smaller doses (0.5 and 1 mg/day), however the maximum dose (2 mg/day) does not seem to be affected by the delivery methods.

Hypothesis: higher doses are more effective

We run test to see if higher doses of Vitamin C are more effective than lower ones in terms of teeth growth. For these ends, we group the records only by the dose levels (0.5, 1, and 2 mg/day) without differentiation by the types of supplement, and compare these three groups to each other. We use **dplyr** library for convenience.

```
library(dplyr)
dose2 <- filter(ToothGrowth, dose == 2)
dose1 <- filter(ToothGrowth, dose == 1)
dose0.5 <- filter(ToothGrowth, dose == 0.5)
t.test(dose2$len, dose1$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  dose2$len and dose1$len
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387     Inf
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

```
t.test(dose1$len, dose0.5$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1$len and dose0.5$len
## t = 6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

```
t.test(dose2$len, dose0.5$len, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  dose2$len and dose0.5$len
## t = 11.799, df = 36.883, p-value = 2.199e-14
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  13.27926      Inf
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

Each of three tests estimated if a higher dose is more effective than the lower one (2 mg to 1 mg, 1 mg to 0.5 mg, and 2 mg to 0.5 mg). All the tests resulted in extremely low p-value allowing to reject null hypothesis that the difference is 0 in favour of the alternative hypothesis that the difference is greater than 0. High t-statistic in all the cases (above 1.645 while we use 95% confidence interval) also support this decision.

Hypothesis: maximum dose is not affected by the delivery methods

We run the test to see if there is difference in efficiency for the maximum dose (2 mg) with respect to the type of supplement. Here t-statistic is close to 0, p-value is high, and 95% confidence interval contains 0. We fail to reject null hypothesis, in other words, the delivery methods do not affect teeth growth for pigs receiving 2 mg of Vitamin C per day.

```
t.test(len ~ supp, data = dose2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Hypothesis: orange juice is more effective than ascorbic acid

We run the test to see if orange juice appears to be more effective in the overall experiment without differentiation by the dose levels. The test rejects null hypothesis in favour of the alternative one supporting that orange juice demonstrates higher efficiency. P-value is lower than 0.5, t-statistic is above 1.645, and 95% confidence interval is above 0.

```
t.test(len ~ supp, alternative = "greater", data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Conclusions

We used 95% confidence intervals to evaluate our hypotheses. We found out that higher doses are more effective in general, and the 2 mg dose results are not affected by the type of supplement of Vitamin C. We also found out that orange juice is more effective than ascorbic acid in general, which may be useful if for some reason lower doses are needed; however, it needs additional testing to challenge 0.5 mg and 1 mg groups as well. We assume that the pigs in the study were randomly selected from the larger population of guinea pigs and randomly assigned to treatment groups. Another assumption is that the dose and delivery methods were the only variables to vary across the groups.